

## Forschungsnotiz

Ekkehard Felder/Marcus Müller/Friedemann Vogel

## Das Heidelberger Korpus

### Gesellschaftliche Konflikte im Spiegel der Sprache

0. Überblick
1. Projektziele
2. Gemeinsame Datenstruktur der an HeideKo beteiligten Untersuchungsprojekte
3. Teilprojekte und Teilkorpora des HeideKo
4. Literatur

#### 0. Überblick

*Projektleitung:* Prof. Dr. Ekkehard Felder

*Ort:* Universität Heidelberg

*Aufbauphase:* April 2008 bis Dezember 2011

*Technische Konzeption:* Friedemann Vogel M.A.

*Beteiligte Wissenschaftler:* Prof. Dr. Ekkehard Felder, Dr. Marcus Müller, Dr. Jörn Stegmeier, Friedemann Vogel M.A.

*Studentische Hilfskräfte:* Thomas Bögel, Anja Chaluppa, Johanna Emmerich, Benjamin Frech, Clara Herdeanu, Katharina Jakob, Franziska Köder, Katrin Laier, Anna Mattfeld, Alexandra Nuñez, Cornelia O'Neill, Evi Schedl

Gefördert von der FAZIT-Stiftung und dem Verlag Gruner & Jahr

#### 1. Projektziele

Das Projekt ‚Heidelberger Korpus‘ (HeideKo) entsteht am Lehrstuhl für deutsche Gegenwartssprache (Prof. Felder) als eine gemeinsame Infrastruktur für thematische Korpora, die zu kultur- und diskurslinguistischen Zwecken erstellt werden. Gemeinsames Erkenntnisinteresse der vier Teilprojekte ist die Analyse **gesellschaftlicher Konflikte im Spiegel der Sprache**. Gesellschaftliche Konflikte werden maßgeblich als sprachlich konstituierte Wissenskonflikte („Semantische Kämpfe“; Felder 2006; Felder & Müller 2009) ausgetragen, die wiederum in linguistisch fundierten Diskursanalysen zugänglich sind. Methodisch besteht das gemeinsame Ziel in der Weiterentwicklung diskurslinguistischer Methoden durch eine Kombination aus quantitativen mit qualitativen Analyseverfahren (z.B. Müller i.Dr.). Die Teilprojekte untersuchen dabei erstens **Deutungskonflikte** im Mediendiskurs zu Bau und Fall der Mauer bzw. zur Deutschen Einheit; zweitens

**Wissenskonflikte** in Experten- und Laiendebatten der Bioethik; drittens **Normenkonflikte** in der juristischen und medienöffentlichen Aushandlung von Rechtsnormen am Beispiel des umstrittenen Gesetzgebungsverfahrens zur Online-Durchsuchung. Die zentrale Datenorganisation ermöglicht einzelnen Untersuchungsprojekten zum einen kontrastive Analysen durch Rückgriff auf andere themengebundene Referenzkorpora, zum anderen den Zugriff auf gemeinsame Analyseinstrumente.

Soziale Wissensbestände und Erfahrungen manifestieren sich in einzelnen Text- und Gesprächsexemplaren, die thematisch stets in Geflechte eingebunden sind (Felder/Müller (Hg.) 2009). Diese Text- und Gesprächsgeflechte sind zwar nicht identisch mit dem Diskurs, aber ein analytisch fassbarer Ausschnitt seiner performativen Konstitution (Busse & Teubert 1994). Sie können als diskursive Spuren (Pêcheux 1969) und Indizes (Peirce CP 2.247) für soziale Interaktionen betrachtet werden. Eine solche These verlangt nach einer genauen Untersuchung von Zeichen und Zeichenverkettungen in Diskursen unter Berücksichtigung der sprachwissenschaftlichen Forschungen zur Kontextualisierung (vgl. z.B. Auer 1986, Feilke 1994, Busse 2007) und zur immanenten Perspektivität sprachlicher Zeichen (Köller 2004). Vor diesem Hintergrund ergeben sich im Hinblick auf das eingangs skizzierte Erkenntnisinteresse folgende präzisierende Fragen:

- Welches Erkenntnispotential vermag die Ermittlung von Sprachgebrauchsmustern in diskursiv und thematisch zusammenhängenden Texten und Gesprächen für die Untersuchung gesellschaftlicher Konflikte zu eröffnen (rekurrente Ausdruckskomplexe als Indikatoren für Denkraum und handlungsleitende Konzepte – Baker 2006, Felder 2009, Bubenhofer 2009)? Welcher Zusammenhang besteht zwischen Formulierungspräferenzen und sozialen Rollen der Diskursakteure?
- Wie lässt sich das komplexe Verhältnis zwischen sprachlichen Mitteln auf der Oberfläche von Texten bzw. Gesprächen und den als Interpretationshypothesen zu fassenden Aussagen im Foucaultschen Sinne mit linguistischen Beschreibungskriterien bestimmen (Konerding 2005, Felder 2009, Müller 2007, Vogel 2009)?

Während unter dem englischen Label *discourse analysis* sprachpragmatische Studien sowie solche zur Organisation von Texten und Gesprächen und zur Sprachvariation in verschiedenen Wissens- und Gesellschaftsdomänen bereits recht ausführlich und erfolgreich korpuslinguistisch durchgeführt wurden (z. B. Biber, Connor & Upton 2007, Überblick in Virtanen 2009), steckt die hier adressierte korpuslinguistische Kulturanalyse weitestgehend in den Anfängen (Baker 2006, Baker et al 2008, Bubenhofer 2009, Mauthner 2009).

## 2. Gemeinsame Datenstruktur der an HeideKo beteiligten Untersuchungsprojekte

Arbeitsgrundlage bildet dabei eine hierarchisch strukturierte Access-Datenbank, deren Funktionalität durch VBA- und Python-Programmierung im Hinblick auf

diskurslinguistische Fragestellungen (wie die Visualisierung von intertextuellen Verknüpfungen u.ä.) erweitert wird. Es handelt sich um eine multimediale Datenbank mit den Dateitypen Text, Transkript, Audiofile, Videofile, PDF, Bild. Für jeden Dateityp stehen separate Metadaten-Sets zur Verfügung. Die einzelnen Datensätze und -typen lassen sich untereinander in Beziehung setzen. Hierdurch können systematisch intersemiotische Zusammenhänge dokumentiert und für eine spätere Analyse herangezogen werden. Der unterschiedlichen Performanz diskursiver Ereignisse wird damit Rechnung getragen.

Für die nähere Arbeit mit dem HeideKo wurden diverse Import-, Bearbeitungs- und Exportformulare entwickelt, so dass auf eine Arbeit mit Kommandozeilen weitestgehend verzichtet und HeideKo auch in die Lehre integriert werden kann. Die Exportformulare ermöglichen dabei Filterung, Vorsortierung auf Basis der Metadaten und einfachen Export der Datenbank-Daten in externe Analysesoftware (wie AntConc; vgl. L. Anthony: 2005) und Formate (TXT, XML). Benutzerfreundlichkeit, umfassende Dokumentation und Tagsets der Korpora garantieren darüber hinaus den nachhaltigen Aufbau sowie die Kompatibilität der HeideKo-Daten.

Die Korpusprimärdaten werden umfassend digitalisiert und in Rohform in die Datenbank eingespeist. In individuellen Teilkorpora können beim Export automatisch Wortarten annotiert werden (POS-Tagging). Die mündlichen Sprachdaten werden orthographisch transkribiert.<sup>1</sup>

Die am HeideKo beteiligten Teilkorpora werden mit gleichen Basismetadaten sowie projektspezifischen Metadaten ausgezeichnet. Neben den Basismetadaten zu Medium, Erscheinungsdatum usw. werden Text-, Gesprächs- bzw. Bildsorte, Oberthema, Redegegenstand, situatives Setting sowie akteursbezogene Daten wie Akteursberuf und -institution, Sprecherrolle und soziale Rolle vermerkt.<sup>2</sup> Die sehr genaue und aufwändige Auszeichnung von Sprecherdaten erlaubt die Analyse von Korrelationen von sprachlichen und sozialen Mustern und ist damit die Kerninformation zur diskurslinguistischen Arbeit.

Zugriff auf Teilkorpora von HeideKo haben in Folge laufender urheberrechtlicher Verhandlungen derzeit nur unmittelbar Forschungsbeteiligte sowie in Teilen Studierende während entsprechender Lehrveranstaltungen.

### 3. Teilprojekte und Teilkorpora des HeideKo

Die **linguistische Mediendiskursanalyse des öffentlichen Diskurses zu Bau und Fall der Berliner Mauer** fokussiert dabei die Frage, wie medial konstituierte Sachverhalte und Diskursobjekte als Teil historischer Deutungskonflikte sprach-

1 Als Abweichung von der orthographischen Norm gilt (wie z.B. in GAT) die generelle Kleinschreibung. Markierte Betonungen werden mit Großschreibung gekennzeichnet. Markierte para- und nonverbale Performanzen werden nach GAT in filterbaren Klammerausdrücken annotiert.

2 Vgl. zur diskurspragmatisch orientierten Korpuslinguistik Felder/Müller/Vogel (i. V.).

lich je zeitspezifisch perspektiviert werden und mit welchen linguistischen Beschreibungsmethoden diese zeitgeschichtlichen Perspektivierungsunterschiede expliziert werden können (Ansprechpartner: Ekkehard Felder). Die zu diesem Zweck erhobenen Texte und Visualisierungsformen aus überregionalen seriösen Zeitungen und Zeitschriften umfassen einen Zeitraum von 1961 bis 2005. Derzeit sind bereits rund 3047 Texte digitalisiert und 2,1 Millionen laufende Wortformen korpuslinguistisch bearbeitbar. Die Gesamtgröße aller digitalisierten Medientexte beläuft sich auf voraussichtlich 7,5 Mio. laufende Wortformen.

Das Teilkorpus Bioethik dient als empirische Basis eines **Forschungsprojektes zu divergenten Kontextualisierungen in Mündlichkeit und Schriftlichkeit am Beispiel der Bioethikdebatte**. (Ansprechpartner: Marcus Müller) Die Untersuchung fokussiert die diskursstrukturellen Hintergründe mündlicher und schriftlicher Interaktionen, in denen Verstehensprobleme indiziert sind, ohne metakommunikativ verhandelt zu werden. Daraus erwächst die Aufgabe, interdisziplinäre Debatten als Schnittstellen divergierender Kontexte zu beschreiben. Als Geltungsrahmen aktueller Kontextualisierungspotenziale werden thematische Kontexte (z. B. ‚Transplantationsmedizin‘), funktionale Kontexte (z. B. ‚juridischer Diskurs‘) und situative Kontexte (z. B. ‚parlamentarische Anhörung‘) analysiert (vgl. Müller i. Dr.). Dieses Forschungsinteresse erfordert eine (im korpuslinguistischen Sinne) mittelgroße Menge an thematisch, situativ und funktional identifizierten Sprachdaten. Im Zeitraum eines Jahres wurden mündliche und schriftliche Sprachdaten aus Äußerungen zur Bioethik-Debatte<sup>3</sup> erhoben. Insgesamt hat das Korpus eine Größe von 5 Mio. Tokens und umfasst unterschiedliche kommunikative Gattungen (mündlich: u.a. Diskussion, Interview, Call-In; schriftlich: v.a. Internetforum, policy paper, Debattenbeitrag u.ä.; Wortlautprotokolle: Debatten im deutschen Ethikrat, Internetprotokolle älterer Fernsehsendungen). – Thematisch und funktional mit dem Teilkorpus ‚Bioethik‘ überschneidet sich das Teilprojekt zur **diskurslinguistischen Analyse der semantischen Kämpfe um Palliativmedizin bzw. Sterbehilfe** (Ansprechpartner Jörn Stegmeier, Hubert Bardenheuer<sup>4</sup> und Ekkehard Felder). Es verfolgt das gleiche Erkenntnisinteresse wie das Mauer-Projekt, fokussiert jedoch weniger auf diachrone, als vielmehr synchrone (domänenvergleichende) Analysen.

Die Untersuchung von **Rechtsnormgenese Prozessen im Gesetzgebungsverfahren am Beispiel der ‚Online-Durchsuchung‘** (Ansprechpartner: Friedemann Vogel) lenkt den Blick schließlich auf juristische Normierungskonflikte im Spiegel der (Medien-)Öffentlichkeit. Rechtsnormgenese stellt sich dabei dar als

3 ‚Bioethik‘ steht hier als Oberbegriff für die thematischen Kontexte ‚Rote Gentechnologie‘, ‚Grüne Gentechnologie‘, ‚Stammzellforschung‘, ‚Reproduktionsmedizin‘, ‚Lebensbeginn‘, ‚Mensch im Gesundheitssystem‘, ‚Transplantationsmedizin‘, ‚Sterbehilfe‘, ‚Pflege‘, ‚Inzest‘, ‚Enhancement‘, ‚Allgemeine Bioethik‘.

4 Chefarzt der Universitären Palliativstation am Krankenhaus St. Vincentius und Universitätsprofessor der Medizinischen Fakultät in Heidelberg.

ein Prozess „reflexiver Transkriptivität“ (Jäger 2003), in dem unterschiedlichste Diskursakteure die ihnen jeweils zur Verfügung stehenden sprachlichen Mittel ko(n)textuell dazu einsetzen, die umstrittenen Normtexte in den verschiedenen Diskursräumen aus ihrer Perspektive konzeptuell dominant zu setzen (vgl. auch Müller 1995). Ziel des Projektes ist der diachrone Nachvollzug der diskursiven Auseinandersetzung in den verschiedenen institutionellen und kommunikativen Räumen und die Konstitutionsbedingungen juristischer Normgenese.

Um dem breiten Diskursraum juristischer Aushandlungsprozesse gerecht zu werden, wird ein umfassendes juristisches Korpus rund um das Thema der „On-line-Durchsuchung“ zusammengestellt aus Texten der Legislative und Exekutive (Gesetze und Gesetzesentwürfe, Parlaments- und Ausschussprotokolle, Stellungnahmen der Parteien und Verbände usw.), der Judikative (insb. Urteile von BGH und BVerfG), des rechtswissenschaftlichen Binnendiskurses (Kommentare, wissenschaftliche Artikel u.ä.), sowie der Medienberichterstattung (Zeitungsberichte, -meldungen, -kommentare; Online-Foren- und Blog-Einträge). Das Korpus umfasst derzeit rund 4 Mio. laufende Wortformen.

#### 4. Literatur

- Anthony, Laurence (2004): AntConc: A Learner and Classroom Friendly, Multi-Platform Corpus Analysis Toolkit. In: Proceedings of IWLeL 2004: An Interactive Workshop on Language e-Learning, pp. 7–13.
- Auer, Peter (1986): Kontextualisierung. In: *Studium Linguistik* 19, S. 22–47.
- Baker, Paul (2006): *Using Corpora in Discourse Analysis*. London.
- Baker, Paul/Gabrielatos, Costas/Khosravini, Majid/Krzyzanowski, Michal/Mcenery, Toni and Wodak, Ruth (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. In *Discourse & Society*, 19/3, S. 273–306. [WWW-document] <http://das.sagepub.com/cgi/content/abstract/19/3/273> (13.03.2009)
- Biber, Douglas/Connor, Ulla and Upton, Thomas A. (2007) *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure*. Amsterdam.
- Bubenhof, Noah (2009): *Sprachegebrauchsmuster. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse*. Berlin/New York (Sprache und Wissen Bd. 4).
- Busse, Dietrich/Teubert, Wolfgang (1994): Ist Diskurs ein sprachwissenschaftliches Objekt? Zur Methodenfrage der historischen Semantik. In: Busse, Dietrich/Hermanns, Fritz/Teubert, Wolfgang (Hg.): *Begriffsgeschichte und Diskursgeschichte. Methodenfragen und Forschungsergebnisse der historischen Semantik*. Opladen, S. 10–28.
- Busse, Dietrich (2007): Diskurslinguistik als Kontextualisierung – Sprachwissenschaftliche Überlegungen zur Analyse gesellschaftlichen Wissens. In: Warnke, Ingo (Hg.): *Diskurslinguistik nach Foucault. Theorie und Gegenstände*. Berlin/New York, S. 81–105 (Linguistik – Impulse und Tendenzen 25).
- Feilke, Helmuth (1994): *Common sense-Kompetenz. Überlegungen zu einer Theorie des „sympathischen“ und „natürlichen“ Meinens und Verstehens*. Frankfurt.
- Felder, Ekkehard (2006): *Semantische Kämpfe. Macht und Sprache in den Wissenschaften*. Berlin/New York.

- Felder, Ekkehard (2009): Sprache – das Tor zur Welt!? Perspektiven und Tendenzen in sprachlichen Äußerungen. In: Felder, Ekkehard (Hg.): Sprache. Im Auftrag der Universitätsgesellschaft Heidelberg. Berlin u.a.: Springer Verlag, S. 13–57 (Heidelberger Jahrbücher Band 53).
- Felder, Ekkehard/Müller, Marcus (Hg.) (2009): Wissen durch Sprache. Theorie, Praxis und Erkenntnisinteresse des Forschungsnetzwerks „Sprache und Wissen“. Berlin/New York (Sprache und Wissen Bd. 3).
- Felder, Ekkehard/Müller, Marcus/Vogel, Friedemann (Hgg.) (i. V.): Korpuspragmatik. Thematische Korpora als Basis diskurslinguistischer Analysen von Texten und Gesprächen. Berlin/New York.
- Jäger, Ludwig (2003): Transkription – zu einem medialen Verfahren an den Schnittstellen des kulturellen Gedächtnisses. In TRANS. Internet-Zeitschrift für Kulturwissenschaften (15/2003). [WWW-document]  
[http://www.inst.at/trans/15Nr/06\\_2/jaeger15.htm](http://www.inst.at/trans/15Nr/06_2/jaeger15.htm) (23.04.2009)
- Köller, Wilhelm (2004): Perspektivität und Sprache. Zur Struktur von Objektivierungsformen in Bildern, im Denken und in der Sprache. Berlin/New York.
- Konering, Klaus-Peter (2005): Diskurse, Themen und soziale Topik. In: Fraas, Claudia/Klemm, Michael (Hg.): Mediendiskurse. Frankfurt/Main, S. 9–38.
- Mauthner, Gerlinde (2009): Checks and balances. how corpus linguistics can contribute to CDA. In: Wodak, Ruth/Meyer, Michael (eds.): Methods of Critical Discourse Analysis. London, S. 122–143.
- Müller, Friedrich (1995): Juristische Methodik. Berlin: Duncker & Humblot.
- Müller, Marcus (2007): Geschichte – Kunst – Nation. Die sprachliche Konstituierung einer ‚deutschen‘ Kunstgeschichte aus diskursanalytischer Sicht. Berlin/New York: de Gruyter (Studia Linguistica Germanica 90).
- Müller, Marcus (i. Dr.): Freiheit, Wille, Verantwortung. Zur Rolle dreier Totalitätsbegriffe in der Bioethik-Debatte. In: Constanze Spieß (Hg.), Sprachstrategien und Kommunikationsbarrieren. Zur Rolle und Funktion von Sprache in bioethischen Diskursen. Bremen: Hempen.
- Pêcheux, Michel (1969): Analyse automatique du discours. Paris: Dunod.
- Peirce, Charles S. (1960): Collected Papers. Cambridge, Massachusetts: Belknap Press. [Zitiert als CP.]
- Virtanen, Tuija (2009): Corpora and Discourse Analysis. In: Lüdeling, Anke/Kytö, Merja (eds.): Corpus Linguistics. Berlin/New York, S. 1043–1070.
- Vogel, Friedemann (2009): Aufstand – Revolte – Widerstand. Linguistische Mediendiskursanalyse der Ereignisse in den Pariser Vorstädten 2005. Frankfurt am Main: Peter Lang (Europäische Hochschulschriften. Reihe 21, Linguistik Bd. 343)

*Adressen der Verfasser:*

*Prof. Dr. Ekkehard Felder/Dr. Marcus Müller/Friedemann Vogel M.A., Germanistisches Seminar, Universität Heidelberg, Hauptstraße 207–209, D-69117 Heidelberg.  
 E-Mail: [ekkehard.felder@gs.uni-heidelberg.de](mailto:ekkehard.felder@gs.uni-heidelberg.de)/ [marcus.mueller@gs.uni-heidelberg.de](mailto:marcus.mueller@gs.uni-heidelberg.de)/ [friedemann.vogel@gs.uni-heidelberg.de](mailto:friedemann.vogel@gs.uni-heidelberg.de)*